Exploring Human Mobility Characteristics Based on Floating Car Data and Mobile Phone Records

Yaofei Chang[*], Chun Liu[* **], De Wang[***]

[*] College of Surveying and Geo-informatics, Tongji University, Shanghai, China
[**] Key Laboratory of Advanced Engineering Surveying of NASMG, Shanghai, China
[***] College of Architecture and Urban Plannning, Tongji University, Shanghai, China

Abstract. Along with the promotion of Information and Communication Technology, floating car data and mobile phone records have been extensively used in capturing individuals' movements. In this paper, we discuss corresponding trajectory extraction methods and conduct a spatio-temporal analysis based on two datasets. Conclusively, we discover that the two datasets are both adequate for reflecting human mobility characteristics. Meanwhile, data qualities and inherent properties make them distinguished in depicting human mobility patterns.

Keywords. Floating Car Data, Smartphone-based Positioning, Human Activities, Spatio-temporal Analysis

1. Introduction

Human mobility characteristics have aroused significant interests recently with rapid development of mobile positioning technology and big data methods. What lies behind spatio-temporal activities of metropolitan residents contributes a lot to interpret urban structures (Liu X et al. 2013), intelligent transportation (Ma et al. 2015), land utilization (Pei et al. 2014), tourism management (Zheng et al. 2011), and environment problems (Zheng et al. 2013). Looking back upon those days when tracking of human activities relied on questionnaires and travel diaries, we are fortunate

Published in "Proceedings of the 13th International Conference on Location-Based Services", edited by Georg Gartner and Haosheng Huang, LBS 2016, 14-16 November 2016, Vienna, Austria.

enough to enjoy newly developed methods in big data era, which outperform the traditional ones both in velocity and volume. Commonly used approaches include vehicle-mounted GPS (Liu Y et al. 2012a), smartphone-based positioning (Diao et al. 2015), smart cards (Kim et al. 2014) and check-ins (Cheng et al. 2011).

Born with distinctive advantages, different techniques reflect human mobility characteristics from different aspects. For instance, mobile phone records cover the entire population, which makes it persuasive to depict human mobility rules of the whole city. Vehicle-based GPS is less universal but of high spatial resolution, and thus enables specific hotspots extraction of human activities.

On the other hand, inconsistent data qualities of different techniques bring challenges on modelling and visualization. No matter actively or passively the data is collected, spatio-temporal information is stored with a certain time interval. Therefore, it's necessary to build suitable models to illustrate trajectory patterns concealed in inconsecutive data. Kang (2013) points out that commonly used origin and destination (OD) in trajectory analysis of floating car data (FCD) is a representative of the trip-based method, while mobile phone records follow the sampling-based framework.

In this paper, we focus on revealing human mobility characteristics using FCD and mobile phone records. We are trying to answer two questions: (1) are FCD and mobile phone records adequate for reflecting human mobility characteristics; (2) are there any similarity and divergence between processing methods of these two kinds of data. This paper is organized as follows. *Section 2* introduces the two datasets and related trajectory extraction methods. *Section 3* gives detailed results of temporal and spatial analysis. Then, *Section 4* discusses several issues along with potential applications of the two datasets in human mobility exploration. At last, *Section 5* summarizes the contribution of this paper and offers an outlook of the future work.

2. Data Preprocessing

2.1. Data Description

In this paper, we use two datasets D1 and D2, both gathered in Shanghai, China. D1 is a trajectory dataset collected by 8,000 taxis (equipped with GPS) on seven consecutive days with a time interval of 10 seconds theoretically, adding up to 40 million pieces of data. Each piece of data records the taxi's id, speed, longitude, latitude, height, state, and time. To be specific, "id" identifies a taxi exclusively, location information is based on WGS1984, and “state” indicates the taxi’s occupancy condition, where “0” means available while “1” means occupied.

Since that the full dataset of mobile phone records is excessively big, we extract a uniform sample from one day’s dataset as D2, which contains about 20 million pieces of records and covers approximately seven million residents, accounting for 30% of the permanent population. D2 logs correspondences between base stations and mobile phones, whose time interval is determined by the operating and moving frequency of phone subscribers. Similarly, each record includes an anonymous id related to the particular mobile phone, thus protecting subscribers’ privacy. And location information is present in latitude and longitude as well.

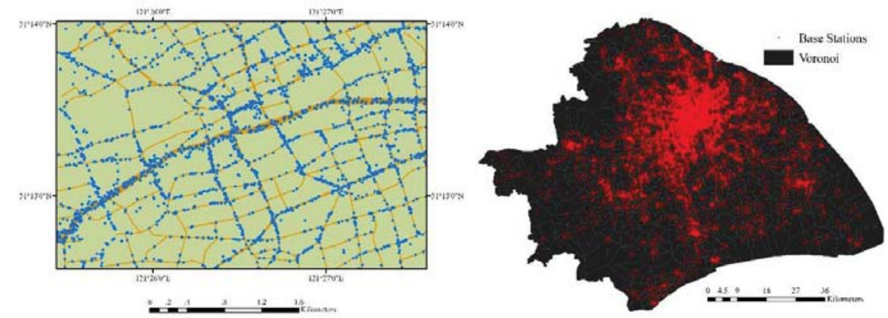


Figure 1. FCD (left) and mobile phone record (right) have different spatial resolution. The former is located on streets while the latter is limited to the range of base stations.

It should be noted that spatial resolution of FCD and mobile phone records are prominently different due to their positioning methods. As *Figure 1* indicates, GPS limits the positioning accuracy to road level, while that of mobile phone records is up to the distribution density of base stations which complies with a diminishing pattern from the urban centers to the surrounding areas.

2.2. Trajectory Extraction

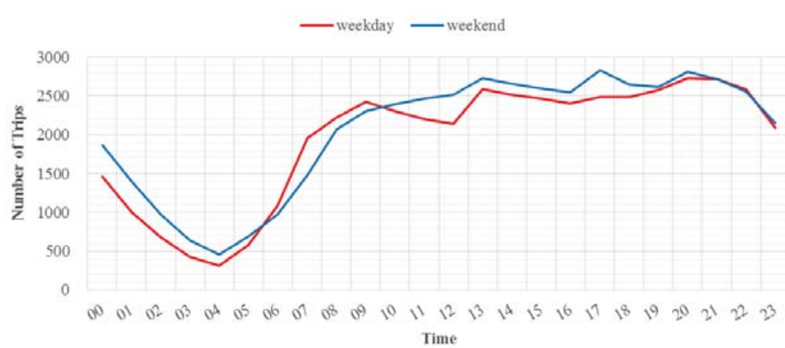
Each record in D1 and D2 can be abstracted by a 3D point feature (x_i, y_i, t_i), where x_i and y_i denote the location, and t_i denotes the time. As for D1, field “state” is a good marker in extracting trajectories. State transformation between “available” and “occupied” helps identifying displacements of metropolitan residents, providing a better understanding of human activity characteristics in the flow space as a result.

When it comes to D2, things are different. Indicators of human activity trajectories seem ambiguous here because mobile phones are not designed to record spatial movements. Hence, sampling-based trajectory extraction is applied to D2 in practical situation, where datasets of one day and peak hours are commonly used exemplars. It's reasonable that most trajectories are closed because home is a stagnation point in ordinary people's daily routine. And we can tell different human activities such as working and staying in with the help of POIs.

3. Human Mobility Characteristics

3.1. Temporal Distributions

It's apparent that time serves as a contributory factor in human mobility. To make a specific understanding, we fetch valid trips (no less than 500 meters) from D1 and display the hourly statistical result in *Figure 2a*. It can be concluded that temporal characteristics of weekends are different from that of weekdays. The morning peak disappears on weekends when most people are free from commuting restrictions. In addition, there is a summit lasting from 17:00 to 18:00 on weekends, which may result from the prologue of night activities. In general, human activities are put off on weekends along with an extended duration, and the traffic volume based on taxis surpasses that on weekdays. Similar patterns have been discovered by Liu (2012b) using FCD in Shanghai collected in 2009, and we are about to investigate the temporal distribution characteristics even more elaborately by shortening the time interval to half an hour or ten minutes.



(a)

(b)

Figure 2. Temporal distribution patterns of (a) FCD and (b) mobile phone records. Three conspicuous peaks of human mobility are shown in (a), both on weekday and on weekends. While (b) indicates that 10:00-11:00 is the busiest period of human activities.

The temporal distribution pattern we draw from D2 exhibits a parallel tendency with that of D1. Nevertheless, it should be pointed out that no midday peak exists in *Figure 2b*. A possible explanation to this phenomenon is that communication acts tend to happen in equal probability during daytime, which makes subscribers' movements the primary factor influencing temporal distribution characteristics of D2. Since that trips within the range of one base station will not be recorded, it's conspicuous why the noon, when most people take a rest or wander around, turns out less active during daytime.

3.2. Spatial Distributions

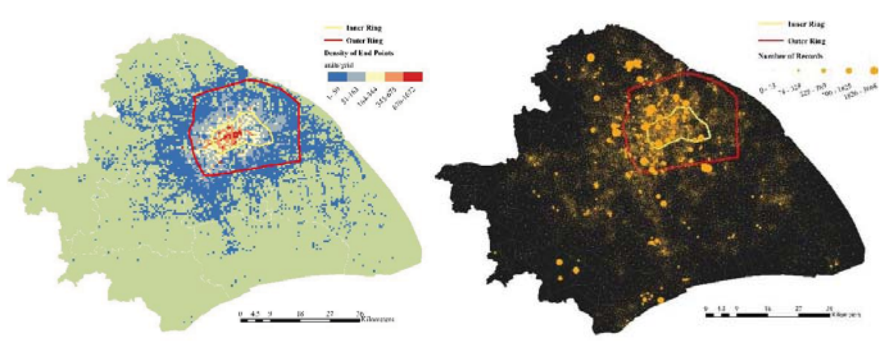


Figure 3. Spatial distribution patterns of FCD (left) and mobile phone records (right). Most frequently visited spots detected by FCD cluster inside the inner ring, while hot spots inferred by mobile phone usages are comparatively discrete in space.

As a peculiar transportation means, taxi usages have strong relationship with the latent urban structure. Apart from particular functional regions like airports and railway stations, taxi usages are most likely to gather in urban areas, especially commercial districts and tourist attractions. What we distinguish from *Figure 3a*, which depicts the spatial distribution of all the get-off points in D1, highly coincides with the hypothesis. Areas in red, which indicates a higher density of get-off points, are clustering within the inner ring of Shanghai, except for Hongqiao International Airport and South Railway Station.

Remarkably, *Figure 3b* demonstrates another spatial pattern of human mobility revealed by mobile phone usages. Noticing that mobile phone usages reflect residents' daily activities instead of certain travel purposes, hotspots detected from D2 are comparatively decentralized in geographical space. Regions with diverse functions may get involved in daily movements, not limited to home, offices, and entertainment places. Therefore, several residential districts resting between the inner and outer rings become hotspots, but typical commercial districts such as People's Square and Lujiazui are not included. And those base stations outside the outer ring with strong connection intensity may either be suburban centers or result from the edge effect.

4. Discussions

Big data, as a newly developed sensor of the real world, enables us to perceive spatial distributions as well as spatial interactions. What we have discussed above demonstrates the merits of FCD and mobile phone records on quantifying spatial distributions of human mobility. Likewise, check-in data from social media also contributes to solve urban issues, such as discovering unreasonable distribution of traffic arteries or detecting highly urbanized regions nationwide. Furthermore, thanks to the capability of revealing interpersonal connections, big data makes it possible to quantify interdependence and interactive intensity of geographical units. To be specific, researches on interdependence excel at deriving regions with particular attributes on a rough scale, thus commonly employed in land use classification. While detecting detailed urban structures is the expertise of interactive intensity, where the community discovery algorithm is generally used.

Before the advent of big data, small data collected by conversations and questionnaires occupied an indispensable position in geographical researches. Small data is rich in individual information, like age, sex, family relationships and penchants. However, big data shows drawbacks in what small data is good at. Poor in property information restricts big data to designated themes. For example, FCD only reflects a part of individuals' movements and fails to distinguish their travel purposes, so it will be regarded less valuable in researches of travel requirements of urban residents. On the other hand, quantity of big data may become the stumbling block to the quality of output. In other words, abundant data does not necessarily lead to accurate conclusions, which has been discussed thoroughly by Zhao (2016). Therefore, the emergence of big data brings about a people-oriented approach into geographical researches along with potential challenges on data processing and discrimination. What we should pay attention to when excavating into big data is that the balance between quality and quantity is always a vital point. Only appropriate data can result in convincing discoveries.

5. Conclusions

In this paper, we generate a comparative analysis of human mobility characteristics in Shanghai based on floating car data and mobile phone records. In general, we find the two datasets both adequate for reflecting human mobility characteristics. However, it should be noted that FCD, as a frequently used means of transportation, is better at detailed trajectory illustration. And it follows a concentrated distribution pattern in geographical space and is sensitive to changes temporally. As for mobile phone records, where the sampling-based method is applied to extract trajectories, bring an unparalleled performance in detecting workplaces and residences.

Still, many potential works relied on FCD and smartphone-based positioning can be promoted in the future. An integrated excavation into various datasets will broaden our understanding of geographical interactions and heterogeneity in human mobility patterns. Thus, dynamic monitor and prediction of human activities across various scales will become a promising direction as well as uncovering connections among different functional regions.

Acknowledgement

Yaofei Chang would like to thank Professor De Wang for his constructive suggestions and support.

References

Cheng Z, Caverlee J, Lee K, et al. Exploring Millions of Footprints in Location Sharing Services[C]. International Conference on Weblogs and Social Media, 2011.

Kang C, Sobolevsky S, Liu Y, et al. Exploring human movements in Singapore (a comparative analysis based on mobile phone and taxicab usages)[J]., 2013.

Kim K, Oh K, Lee Y K, et al. An analysis on movement patterns between zones using smart card data in subway networks[J]. International Journal of Geographical Information Science, 2014, 28(9): 1781-1801.

Liu X, Gong L, Gong Y, et al. Revealing daily travel patterns and city structure with taxi trip data[J]. Journal of Transport Geography, 2013, 43(0).

Liu Y, Wang F, Xiao Y, et al. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai[J]. Landscape and Urban Planning, 2012a, 106(1): 73-87.

Liu Y, Kang C, Gao S, et al. Understanding intra-urban trip patterns from taxi trajectory data[J]. Journal of Geographical Systems, 2012b, 14(4): 463-483.

Ma S, Zheng Y, Wolfson O, et al. Real-Time City-Scale Taxi Ridesharing[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(7): 1782-1795.

Diao M, Zhu Y, Ferreira J, et al. Inferring individual daily activities from mobile phone traces: A Boston example[J]. Environment and Planning B-planning & Design, 2015.

Pei T, Sobolevsky S, Ratti C, et al. A new insight into land use classification based on aggregated mobile phone data[J]. International Journal of Geographical Information Science, 2014, 28(9): 1988-2007.

Zhao Z, Shaw S L, Xu Y, et al. Understanding the bias of call detail records in human mobility research[J]. International Journal of Geographical Information Science, 2016: 1-25.

Zheng Y, Zhang L, Ma Z, et al. Recommending friends and locations based on individual location history[J]. ACM Transactions on The Web, 2011, 5(1).

Zheng Y, Liu F, Hsieh H, et al. U-Air: when urban air quality inference meets big data[C]. Knowledge Discovery and Data Mining, 2013.